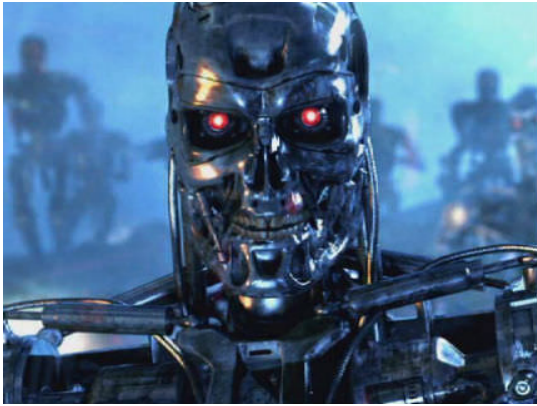


# Let Slip the Robots of War

**Lethal autonomous weapon systems might be more moral than human soldiers.**

Ronald Bailey | Jan. 23, 2015 12:01 am



Terminator

Lethal autonomous weapons systems that can select and engage targets do not yet exist, but they are being developed. Are the ethical and legal problems that such "killer robots" pose so fraught that their development must be banned?

Human Rights Watch thinks so. In its 2012 report, *Losing Humanity: The Case Against Killer Robots*, the activist group demanded that the nations of the world "prohibit the development, production, and use of fully autonomous weapons through an international legally binding instrument." Similarly,

the robotics and ethics specialists who founded the International Committee on Robot Arms Control wants "a legally binding treaty to prohibit the development, testing, production and use of autonomous weapon systems in all circumstances." Several international organizations have launched the Campaign to Stop Killer Robots to push for such a global ban, and multilateral meeting under the Convention on Conventional Weapons was held in Geneva, Switzerland last year to debate the technical, ethical, and legal implications of autonomous weapons. The group is scheduled to meet again in April 2015.

At first blush, it might seem only sensible to ban remorseless automated killing machines. Who wants to encounter the Terminator on the battlefield? Proponents of a ban offer four big arguments. The first is that it is just morally wrong to delegate life and death decisions to machines. The second is that it will simply be impossible to instill fundamental legal and ethical principles into machines in such a way as to comply adequately with the laws of war. The third is that autonomous weapons cannot be held morally accountable for their actions. And the fourth is that, since deploying killer robots removes human soldiers from risk and reduces harm to civilians, they make war more likely.

To these objections, law professors Kenneth Anderson of American University and Matthew Waxman of Columbia respond that an outright ban "trades whatever risks autonomous

weapon systems might pose in war for the real, if less visible, risk of failing to develop forms of automation that might make the use of force more precise and less harmful for civilians caught near it."

Choosing whether to kill a human being is the archetype of a moral decision. When deciding whether to pull the trigger, a soldier consults his conscience and moral precepts; a robot has no conscience or moral instincts. But does that really matter? "Moral" decision-making by machines will also occur in non-lethal contexts. Self-driving cars will have to choose what courses of action to take when a collision is imminent—e.g., to protect their occupants or to minimize all casualties. But deploying autonomous vehicles could reduce the carnage of traffic accidents by as much as 90 percent. That seems like a significant moral and practical benefit.

"What matters morally is the ability consistently to behave in a certain way and to a specified level of performance," argue Anderson and Waxman. War robots would be no more moral agents than self-driving cars, yet they may well offer significant benefits, such as better protecting civilians stuck in and around battle zones.

But can killer robots be expected to obey fundamental legal and ethical principles as well as human soldiers do? The Georgia Tech roboticist Ronald Arkin turns this issue on its head, arguing that lethal autonomous weapon systems "will potentially be capable of performing more ethically on the battlefield than are human soldiers." While human soldiers are moral agents possessed of consciences, they are also flawed people engaged in the most intense and unforgiving forms of aggression. Under the pressure of battle, fear, panic, rage, and vengeance can overwhelm the moral sensibilities of soldiers; the result, all too often, is an atrocity.

Now consider warbots. Since self-preservation would not be their foremost drive, they would refrain from firing in uncertain situations. Not burdened with emotions, autonomous weapons would avoid the moral snares of anger and frustration. They could objectively weigh information and avoid confirmation bias when making targeting and firing decisions. They could also evaluate information much faster and from more sources than human soldiers before responding with lethal force. And battlefield robots could impartially monitor and report the ethical behavior of all parties on the battlefield.

The baseline decision-making standards instilled into war robots, Anderson and Waxman suggest, should be derived from the customary principles of distinction and proportionality. Lethal battlefield bots must be able to make distinctions between combatants and civilians and between military and civilian property at least as well as human soldiers do. And the harm to civilians must not be excessive relative to the expected military gain. Anderson and Waxman acknowledge that current robot systems are very far from being able to make such

judgments reliably, but do not see any fundamental barriers that would prevent such capacities from being developed incrementally.

Individual soldiers can be held responsible for war crimes they commit, but who would be accountable for the similar acts executed by robots? University of Virginia ethicist Deborah Johnson and Royal Netherlands Academy of Arts and Sciences philosopher Merel Noorman make the salient point that "it is far from clear that pressures of competitive warfare will lead humans to put robots they cannot control into the battlefield without human oversight. And, if there is human oversight, there is human control and responsibility." The robots' designers would set constraints on what they could do, instill norms and rules to guide their actions, and verify that they exhibit predictable and reliable behavior.

"Delegation of responsibility to human and non-human components is a sociotechnical design choice, not an inevitable outcome technological development," Johnson and Noorman note. "Robots for which no human actor can be held responsible are poorly designed sociotechnical systems." Rather than focus on individual responsibility for the robots' activities, Anderson and Waxman point out that traditionally each side in a conflict has been held collectively responsible for observing the laws of war. Ultimately, robots don't kill people; people kill people.

Would the creation of phalanxes of war robots make the choice to go to war too easy? Anderson and Waxman tartly counter that to the extent that banning warbots potentially better at protecting civilians for that reason "morally amounts to holding those endangered humans as hostages, mere means to pressure political leaders." The roots of war are much deeper than the mere availability of more capable weapons.

Instead of a comprehensive treaty, Waxman and Johnson urge countries, especially the United States, to eschew secrecy and be open about their war robot development plans and progress. Lethal autonomous weapon systems are being developed incrementally, which gives humanity time to understand better their benefits and costs.

Treaties banning some extremely indiscriminate weapons—poison gas, landmines, cluster bombs—have had some success. But autonomous weapon systems would not necessarily be like those crude weapons; they could be far more discriminating and precise in their target selection and engagement than even human soldiers. A preemptive ban risks being a tragic moral failure rather than an ethical triumph.

*Disclosure: I have made small donations to Human Rights Watch from time to time.*